高技术通讯 2022 年 第 32 卷 第 12 期:1236-1244

doi:10.3772/j.issn.1002-0470.2022.12.003

降低冗余检测框数量的目标检测算法①

王宪保② 吴梦岚 姚明海

(浙江工业大学信息工程学院 杭州 310023)

摘 要 在机器视觉领域中,很多无锚框检测算法在处理目标密集的图像时会产生冗余边界框的现象,降低了检测精度。针对这种现象,本文借助 RetinaNet 的网络结构提出一种可以降低冗余检测框数量的目标检测方法。首先在特征提取阶段,加入一种新的注意力机制来提高特征的表达能力;然后为了减少正样本中标签错误标定的可能,对选取正样本的位置进行筛选,之后将算法选择的正样本输入预测分支得到目标边界框的坐标和置信度;最后根据目标边界框的位置和分类结果,提出一种类内的交并比分数重分配推理策略,该策略能够减少重叠的检测框数量,从而提高算法精度。本文算法的有效性在公开的图像数据集上进行验证,结果表明,所提出的算法可以提高检测精度、优化定位效果,具有较好的应用前景。

关键词 无锚框;目标检测;注意力机制;分数重分配;推理策略

0 引 言

目标检测是目前计算机视觉领域非常活跃的研究方向[1],广泛地应用在自动驾驶、工业检测[2]和视频监控等多个领域。随着深度卷积神经网络(deep convolutional neural network, DCNN)的发展,基于深度学习的目标检测方法具有特征自动提取、泛化能力强的优点,已然成为目标检测的主流方法。基于深度学习的目标检测算法根据是否需要特意生成候选区域可以分为 2 类:单阶段算法和两阶段算法。这 2 种方法都需要对象候选区域来对目标进行回归和分类。为了减少候选区域生成的时间损耗,Ren 等人[3]在更快速的基于区域生成的卷积神经网络(faster region-based convolutional neural network, Faster RCNN)中提出了一种基于锚框的检测方法(anchor-based object detection, ABOD)。该方法用一组形状、数量及变化比例固定的锚框来代替需要算法产生或预划分的对象候选区域。这一方法的提出,大幅提高了单阶段和两阶段算法的精度和速度。但是 ABOD 算法仍存在诸多不足:(1)锚框的形状、数量和变化比例对检测的精度影响很大。(2)为了得到高的召回率,锚框需要尽可能地遍布图像,因此也产生了更多的负样本。(3)算法会涉及到多种与锚框相关的计算,如在迭代训练中不断调整锚框位置的坐标和计算锚框与真值框的交并比(intersection over union, IoU)。

针对 ABOD 算法存在的问题,同时为了计算简便,研究者将越来越多的目光投向了单阶段的无锚框检测算法(anchor-free object detection, AFOD)。Law 等人[4]提出用一组对角点来确定目标对象的边界框位置,从而舍弃算法对锚框的需求。但是这个方法由于要组合对角点,会对最后的检测结果造成很大的不确定性,所以 Duan 等人[5]提出加入对象中心点的检测。上述方法都是先找到目标的关键点,再由关键点定位到目标整体,间接地对目标进行检测。Tian 等人[6]结合全卷积语义分割的思想,提

① 国家自然科学基金(61871350),浙江省科技计划(2019C011123)和浙江省基础公益研究计划(LGG19F030011)资助项目。

② 男,1977 年生,博士,副教授;研究方向:神经网络,机器学习;联系人,E-mail: wxb@zjut.cn。

(收稿日期:2021-05-07)

出全连接单阶段目标检测算法(fully convolutional one-stage object detection,FCOS)算法,该算法直接在最后的特征图上预测目标类别和边界框的位置,实现了像素级别的预测。这种密集检测的方式增加了正样本点的数量,使得前后背景的候选样本数量更加平衡,但是模型的训练效果很容易受样本点的选择方式影响,例如选择远离目标的样本点用于训练,会产生一些低质量的检测框。除此之外,AFOD算法对于低分辨率的图片进行检测时可以得到评估效果较好的结果,但对密集复杂场景的图像则会出现密集重合的目标定位框。

针对上述问题,本文提出了一种降低冗余检测框数量的目标检测方法,算法对密集重合的定位框进行了计算与置信分数分配。本文主要贡献如下。

(1)使用基于锚框的目标检测算法 RetinaNet[7] 的网络结构提出一种可以降低冗余检测框数量的无锚框目标检测算法。

(2)提出了一种增强的通道注意力机制,加入算法中提高其特征表达能力。

(3)提出一种类内分数重分配机制的推理策略,有效抑制了重合框,提高了检测精度。

1 相关工作

本文算法借用了如图 1 所示的 RetinaNet 算法的网络结构,每一层特征图 P_3、P_4、P_5 都是来自于特征提取器的输出 C_3、C_4、C_5 经过卷积和上层特征层上采样相加得到的。对应的 P_6 由 P_5 通过卷积下采样生成,P_7 由 P_6 通过卷积下采样计算生成。整个算法需要主干网络、多尺度特征构建、样本选择和寻找局部最优检测框。

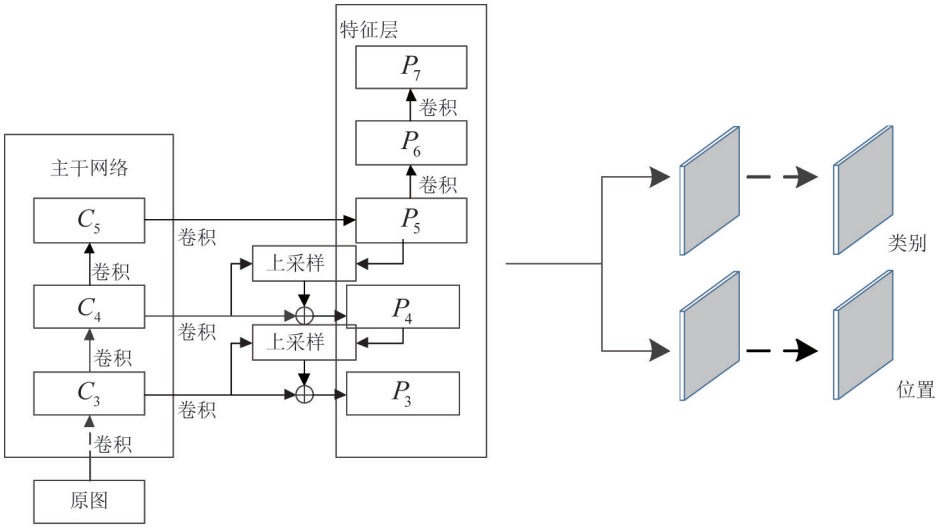

图 1 RetinaNet 的简易结构

1.1 主干网络

VoVNet[8] 与残差网络[9]一样,可以作为模型的主干网络来提取待测图像的特征。VoVNet 网络是由多个一次性聚合(one-shot aggregation, OSA)模块连接而成,OSA 的存在使得模型的计算能力增强,能耗降低。并且由于 OSA 对特征图采用了特殊的连接方式,使得网络可以有效地提取特征。如图 2 所示,OSA 模块中有 2 种连接方式。一种是一组卷积层的连接,这可使网络获得较大的感受野。另一种是在最后一次性地聚集了各个卷积层输出特征,增加了整个网络的特征聚合能力,同时保证输入输出的通道数相同。OSA 的计算公式如式(1)所示。

$$\begin{cases} F_0 = F_{3\times3}(x_i) \\ F_j = F_{3\times3}(F_{j-1}),\ j = 1,\cdots,4 \\ x_{i+1} = F_{1\times1}(x_i \oplus F_0 \oplus F_1 \oplus F_2 \oplus F_3 \oplus F_4) \end{cases} \tag{1}$$

其中,F_j 表示第 j 层卷积层的输出,$F_{3\times3}$ 和 $F_{1\times1}$ 分别代表 3×3 和 1×1 的卷积操作,x_i 和 x_{i+1} 是当前 OSA 模块的输入和输出,$\oplus$ 表示连接计算。

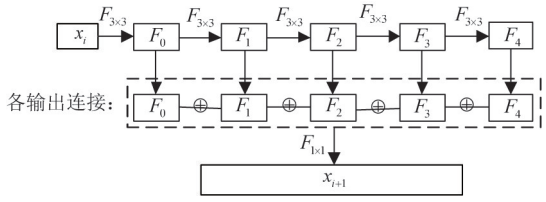

图 2　一次性聚合模块

1.2　多尺度特征

当前的目标检测算法为了检测不同尺度大小的对象，主要将主干网络提取的多尺度特征构成金字塔的形状。所以图像金字塔是一组由原图经过不同的尺寸缩放得到的特征图，从低层到高层，图像的分辨率逐层减少。它在算法的预测中的使用方式主要有：(1)用不同尺度的特征图单独预测不同尺度的目标，如 Wei 等人[10]提出的 SSD(single shot multi-box detector)算法。(2)融合不同尺度的特征图用于预测，如 Lin 等人[11]就此提出了特征金字塔网络(feature pyramid network, FPN)，对主干网络生成的特征图采用自顶向下的方式，将高层的特征向低层传播并与同层特征进行融合，缩小了不同层级的特征图之间的语义差异，增强了特征图的表征能力。

1.3　样本的选择

对于一般的 ABOD 算法，哪个锚框用于回归训练都需要经过计算确定，如 RetinaNet 算法，对于生成的每层特征图像的每个像素点位置都预设多个锚框，然后分别计算每个锚框与真值框的 IoU，如果 IoU 大于设定的阈值，那么这个锚框会被标定为正样本。对应于 AFOD 算法，通过直接寻找关键点或是位置和尺度信息来确定正样本点，如 FoveaBox 算法[12]先将真值框映射到相应特征层上，根据预设位置将真值框内部的特征点 f_i(第 i 层特征层)置为待选正样本，当待选点落在设定的尺度范围之内，那么待选样本点可以被认定为正样本点，它的标签是它所在真值框的标签。这种样本的选择方法与 ABOD 算法相比，可以得到更多的正样本点，前后背景的候选样本点的数量更加均衡，也节省了很多与锚框相关的计算量。Zhang 等人[13]引入了锚框，将样本点等同于锚框中心点，每个样本点对应 8 个锚框。用距离描述锚框中心点与真值框中心点的距离，从中挑选出距离最近的前 k 个锚框，再用锚框与真值框的 IoU 的标准差和平均值的和作为阈值用来挑选符合要求的锚框。如果挑选出来的锚框中心点的位置在真值框内部，那么该中心点是正样本。Qiu 等人[14]利用边界信息来增强有效样本点的信息表达能力。限制样本点的位置，可以减少错误标定的正样本数量；增强样本点的特征表达，能够增强它对应的目标信息，有利于目标回归。

1.4　非极大值抑制

非极大值抑制[15](non maximum suppression, NMS)在很多模型的推理阶段被用来抑制冗余的检测框。算法 1 描述了 NMS 如何利用每个检测框的类别置信度和检测框之间的 IoU 来寻找局部的极大值。

算法 1　NMS 算法

输入：初始检测框的合集 $B = [B_1, \cdots, B_m]$，检测框对应的置信分数集 $S = [S_1, \cdots, S_m]$，IoU 阈值 Tr。

(1) 检测框按照置信分数的大小降序排序，并将置信分数按从小到按从大到小的顺序排列；
(2) 将排序好的置信分数的位置索引值保存到列表 od 中；
(3) While(索引值列表非空)
(4) do{
(5) 将 od 中最大置信分数的位置索引值保存到列表 D 中；
(6) 根据上一步骤的索引值获取置信度最高的检测框 B_{max1}；
(7) 计算 B_{max1} 和其他检测框的 IoU 值，在 od 中删除 B_{max1} IoU 值小于阈值的检测框的位置索引；}
(8) 根据列表 D 中保存的索引值，输出最后的检测框和对应的置信分数。

Liu 等人[16]在多任务的损失函数中加入与类别相关的 NMS 损失，学习每个样本的 NMS 分数，实现类别之间的 NMS 抑制效果。NMS 中加入类别相关的改进，可以得到更好的检测效果，减少目标丢失的可能性和冗余检测框数量。

2　方　法

本文基于 RetinaNet 算法的网络结构提出的目标检测方法与一般的目标检测算法相比，考虑了特征图不同通道的相关性，提出改进 VoVNet。在原 VoVNet 网络提取特征时加入一种增强的改进通道注意力模块(improved squeeze-and-excitation, ISE)，

增强了通道信息的表达,减少了通道信息的丢失。获取图像多层特征之后,进行样本选择,再将样本导入到预测分支,进行算法训练。训练完成之后,在算法推理阶段中,采用提出的类内的交并比分数重分配策略,来减少重合检测框的数量。

2.1 改进 VoVNet

为了提高模型的计算能力,算法使用 VoVNet v1-57 作为特征提取器提取输入图像的特征。为了使特征更好地表达,基于残差网络的启发,在 x_i 与 VoVNet 网络中 OSA 模块输出结合前,加入改进的通道注意力模块。文献[17]中的通道注意力模块(squeeze-and-excitation, SE)使用全局平均池化来挤压特征图通道的空间相关性,然后经过 2 层激活函数分别为 ReLU 和 sigmoid 的全连接层来获取通道的权重。对于特征图 $x \in R^{C\times H\times W}$ 的通道权重 $S(x) \in R^{C\times 1\times 1}$ 的计算过程如式(2)所示。

$$S(x) = \sigma(W_C\delta(W_{C/r}(f_{\text{avg}}(x)))) \tag{2}$$

其中 $f_{\text{avg}}(x)$ 代表全局池化操作, $W_{C/r}$、$W_C \in R^{C\times 1\times 1}$ 是 2 层全连接层的权重,σ 和 δ 分别代表 sigmod 函数和 ReLU 函数。从式(2)中可以看出,SE 中的 2 层特征层连接,势必会对计算造成负担,所以就采取先降维和后升维的操作,减少参数量。让特征图经过第 1 个全连接层从 C 通道减少到 C/r 通道,第 2 个全连接层又将特征图的通道数从 C/r 通道扩充到原通道数。在这通道减少又复原的过程中,会对特征造成通道信息的丢失,所以提出将全连接层减至 1 个,通道数维持为 C 以此提高模块的性能。同时,为了增强全局特征的表达,加入全局最大值池化,计算表达式如式(3)所示,其中 $f_{\max}(x)$ 代表全局最大值池化操作。

$$S(x) = \sigma(W_C(f_{\text{avg}}(x) + f_{\max}(x))) \tag{3}$$

至此,整个 OSA 模块的形式如图 3 所示,其输出如式(4)所示。

$$x'_{i+1} = x_i \otimes S(x_i) + F_{1\times 1}(x_{i+1}) \tag{4}$$

其中 x'_{i+1} 表示输出, x_i 和 x_{i+1} 是原 OSA 模块的输入和输出,⊗表示元素相乘。

2.2 预测

2.2.1 多尺度特征图上的样本点选择

对特征提取器提取的特征融合方式如图1所示,

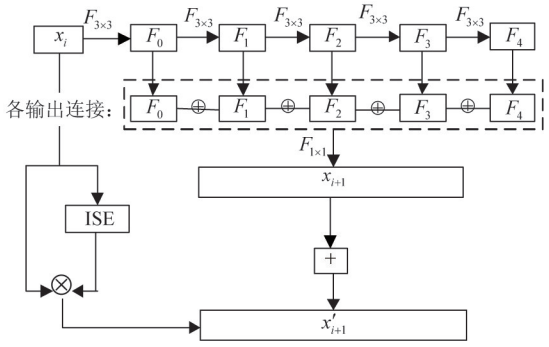

图 3 改进的一次性聚合模块

因此,特征层 $P_3 \sim P_7$ 对输入图片的下采样率为 $\{s_i\} = \{8, 16, 32, 64, 128\}$。

获取到不同级别的特征图之后,就是对特征图上的像素点进行挑选。第 i 层特征层上的点 f_i 映射回原图对应感受野中心的位置坐标为 (x, y), 如果这个点位于真值框内部,且该点到真值框 4 条边的最大垂直距离满足该层预设的尺度范围,就认定点 f_i 是正样本, 其类别标签标定为真值框。考虑到真值框内的点并不都是位于目标对象上:有些样本远离真值框中心,这种样本点预测出来的边界框会偏离被测物体;有些样本是属于背景,会预测出错误的检测框。所以在选择样本点的时候,在真值框内部设置一个正区域:一个与真值框同中心的正方形区域。正区域的设置可以排除一些低质量或是错误标定的样本点,正区域以外的真值框的像素点都不会被输送到损失函数中用于回归分类。在研究中对比发现,与真值框同中心的正方形的边长为 $1.5 \times s_i$ 时,模型的预测效果会好于其他数值。

由于各个级别的特征图尺度不同,可以根据预设的尺寸来决定每层预测对象的大小。在 ABOD 算法中,如 RetinaNet 多尺度特征图 P_3、P_4、P_5、P_6 层对应用于预测回归的锚框面积分别为 64^2、128^2、256^2、512^2。与此相对应,在本文算法中直接对 P_3、P_4、P_5、P_6 层设置的尺度范围为[0, 64]、[64, 128]、[128, 256]、[256, 512], P_7 层设置的尺度范围为[512, ∞],用来限制每个层级预测对象的大小。

2.2.2 预测分支

每层特征图都对应 2 个预测分支,用来预测目标的类别和位置回归。每个预测分支都由 4 层卷积

层构成，分类器最后一层的输出是 20 维的类别标签向量，回归器输出的是一个 4 维的位置向量。训练用的损失函数如式(5)所示。

$$l = \frac{1}{N}\sum_{m,n} Lfl(C_{m,n}, C'_{m,n}) + \frac{1}{N}\sum_{m,n} f(C'_{m,n})Lil(d_{m,n}, d'_{m,n}) \quad (5)$$

式中，N 表示正样本数；m、n 表示的是所有特征图上点对应原图的位置坐标；Lfl 表示的是 focal loss 函数，用于分类训练；$C_{m,n}$ 和 $C'_{m,n}$ 分别表示预测的类别标签和对应的真值框标签；$f(C'_{m,n})$ 是一个非负函数，当 $C'_{m,n} > 0$ 时取 1，$C'_{m,n} \leqslant 0$ 时取 0；Lil 表示的是算法 UnitBox 中的 IoU loss[18] 函数，用于回归训练，其中 $d_{m,n}$ 和 $d'_{m,n}$ 分别表示预测和目标的位置。

至此算法的整体框架如图 4 所示。整体框架包括特征提取、特征层样本点选择和预测分支。

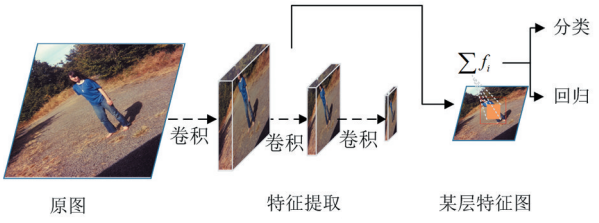

图 4　框架结构图

2.3　推理策略

检测算法对目标分布密集的场景检测容易产生重叠的检测框，如图 5 所示。仅仅采用 NMS 来抑制多余的检测框是不够的，所以在进行 NMS 操作前，加入类内分数重分配和增大类间检测框位置间距 2 个操作。

一张待检测的图片输入到网络中，相当于从特征图上的一组点对目标进行预测，得到一组初步检测的结果，即目标边界框坐标(框的左上角和右下角坐标)、类别标签和类的置信的分数。按照类的置信分数从大到小进行排列，选取前 k 个点的预测结果，按照不同的类别进行处理，如算法 2 所示。对于同一个类别中的检测框 B_c，计算它们之间的 IoU，如果大于某个阈值，则认为它们是同一个目标的检测结果。对同一对象的检测框，根据它们之间的 IoU 进行分数重分配。考虑到再分配的分数与 B_c 中各个检测框之间的 IoU 大小有关，IoU 越大给原来的分增值越多，且不能丢失原有的分数。因此提出检测框分数重分配的规则如式(6)所示。

$$S'_c = \sum_{m=0}^{m} (2I^m - 1) \times S_c \quad (6)$$

式中，S'_c 表示分配后检测框的置信分数，I^m 表示该检测框与第 m 个检测框的 IoU 的值，S_c 是检测框原来的分数。

图 5　AFOD 算法对有密集分布对象的检测结果

算法 2　一种交并比类内重分配机制

输入：某一类的初始检测框 $B_c = [B_{c1}, \cdots, B_{cn}]$，$B_c$ 对应的置信分数 $S_c = [S_{c1}, \cdots, S_{cn}]$，分数阈值 S_tr，IoU 阈值 t。

(1) 获取置信分数大于设定阈值的检测框 B'_c；

(2) 计算 B'_c 中每个检测框之间的 IoU 值；

(3) 当检测框与其他检测框 IoU 的值大于设定的阈值，根据分数重分配规则变更这个检测框的置信分数，当分数分配完成后，获取到置信分数更新过的分数集 S；

(4) 初始检测框 B_c 加上一个与类别有关的偏移来拉大类间的检测框坐标距离以此减少类间检测框重叠现象；偏移大小为类内检测框最大的坐标值与类别索引值的乘积；

(5) 输出用于 NMS 操作的检测框 B_c 及其对应的置信分数 S_c。

所有类的内部分数重分配完成之后，整合之前获取的检测框的分数，再次进行分数排序，执行 NMS 操作。

3 实验结果与分析

本节主要对算法在公共数据集上的测试结果进行对比,以此对算法的有效性作详细说明。首先通过消融实验获取算法需要的参数,然后通过对比本文算法和经典的单阶段无锚框目标检测算法——FCOS 算法在不同主干网络、推理策略的检测精度,证明本文算法检测精度优于 FCOS 算法,加入的通道注意力机制可以提高网络的检测精度,提出的推理策略可以减少冗余检测框数量。

3.1 实验设定

实验硬件平台为 NVIDIA RTX 2060 SUPER 8 GHz、NVIDIA GTX 1060 6 GB,软件平台为 CUDA 10.2、PyTorch 1.7.0 和 Python 3.7。所有实验结果均在上述实验环境中获得。算法使用的数据集是一直被各种检测和分割算法作为训练集和测试集的 Pascal VOC 数据集。由于 Pascal VOC 2012 训练集的图片数量过少,所有的实验都使用 Pascal VOC 2012 的训练集和验证集,共 11 552 张图片作为模型的训练集,用与 Pascal VOC 2012 类别一致的 Pascal VOC 2007[19] 的测试集共 2510 张图片作为模型的测试集。Pascal VOC 收集了 20 种处于不同场景的物体的图片,每张图片都有对应的物体位置、种类和关键点的标签文件。并且数据集中各个类别中的目标在图像中尺度不一,可以满足算法对多尺度目标进行预测的条件。考虑到实验设备的限制,把数据集中用于训练的图片大小全部缩放为 512 像素 ×448 像素。采用带动量的随机梯度下降(stochastic gradient descent,SGD)进行网络的权重更新,其他的实验参数设定除了特定说明都如表 1 所示。

表 1 实验参数设定表

参数	含义	值
batch-size	数据批次大小	16
lr	学习率	0.01
weight-decay	模型权重衰减	0.0001
epochs	训练迭代次数	20
momentum	动量	0.9
NMS-IoU threshold	NMS 操作中 IoU 阈值	0.6
s	类别分数阈值	0.3

3.2 精度比较

本文主要将 FCOS 算法和本文算法训练 20 代后进行对比,通过平均精度(mean average precision,mAP)这一指标来展示不同算法的实验结果。

3.2.1 不同主干网络的精度对比

算法的主干网络分别是 ResNet 50、VoVNet v1-57 和具有残差连接的 VoVNet v1-57,推理策略采用 NMS,它们在 VOC 2007 验证集上的实验结果按照不同的主干网络分成 3 组进行对比,结果如表 2 所示,可以得出以下结论。

(1)比较第 1、2 组算法获得的平均精度(mAP)可以看出,使用 VoVNet v1-57 作为主干网络的算法性能优于使用 ResNet 50 的算法性能,前者网络的检测精度高于后者。

(2)比较第 2、3 组算法的检测结果可知,VoVNet v1-57 网络中加入残差不论对 FCOS 算法还是本文算法,检测的平均精度均有所提高。

(3)比较第 1、2、3 组检测结果可知,本文算法的平均精度均优于 FCOS 算法。

表 2 Pascal 2007 的检测结果

算法	主干网络	*mAP*/%
FCOS	ResNet 50	61.8
本文算法	ResNet 50	64.9
FCOS	VoVNet v1-57	63.7
本文算法	VoVNet v1-57	65.4
FCOS	VoVNet v1-57 + Res	67.3
本文算法	VoVNet v1-57 + Res	67.8

3.2.2 加入注意力机制的算法精度对比

算法的主干网络是在残差中加入通道注意力机制的 VoVNet v1-57,即改进 VoVNet。它们在 VOC 2007 验证集上的实验结果如表 3 所示,其中 + ISE 表示加入通道注意力机制模块。结合表2,从实验

表 3 Pascal 2007 的检测结果

算法	主干网络	*mAP*/%
FCOS	改进 VoVNet	69.7
本文算法	改进 VoVNet	69.9

结果可以看出,在残差部分加入通道注意力机制,与单独引入残差连接的网络相比,其平均检测精度提高幅度更大。

除了平均精度对比,本文还做了主干网络为 VoVNet v1-57 和加入通道注意力机制的 VoVNet v1-57 的训练损失曲线对比,结果如图 6 所示。

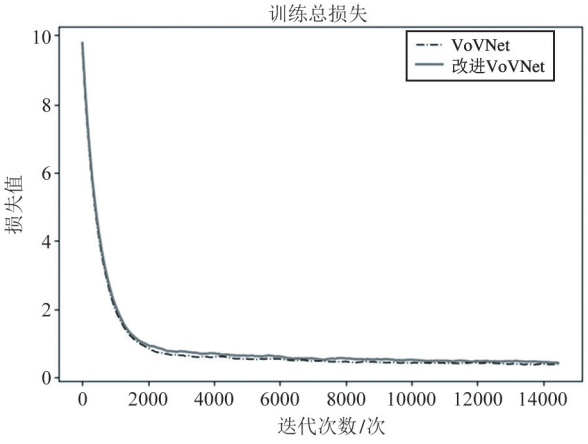

图 6　算法训练的损失曲线

图 6 中虚线表示的是主干网络为 VoVNet v1-57 的整体损失曲线,实线表示的是加入通道注意力机制的 VoVNet v1-57 的整体损失曲线。结合表 2 和表 3 的实验结果,从图中可以看出两个算法的收敛情况基本一致,但是在残差连接处加入通道注意力机制的 VoVNet v1-57 检测效果得到明显提高。

3.2.3　不同推理策略的算法精度和检测框数量对比

对于一个检测目标而言,过分重叠的检测框,与真值框的 IoU 的值都会满足所设定的阈值,除了提高算法的精度,只能从其他方面进行衡量,因此提出用检测框与真值框的数量差 box-d 作为辅助衡量指标。在采用 NMS 和本文推理策略对算法进行验证前,通过消融实验获取算法推理策略中需要的 IoU 阈值,实验结果表 4 所示。

表 4　关于推理策略阈值的实验结果

阈值	0.5	0.6	0.7	0.75	0.8
mAP/%	69.2	69.3	69.3	69.2	69.2

表 4 中的实验是在算法的主干网络为加入残差的 VoVNet v1-57、训练 20 代的情况下做的对比实验。从实验结果来看,对于本文算法,推理策略中使用的 IoU 阈值为 0.6 或是 0.7 时,算法能取得最好的检测结果。因此随后的实验中,本文算法提出的推理策略的 IoU 阈值均设为 0.6。

获取所需的 IoU 阈值后,分别对算法使用 NMS 和提出的推理策略,实验结果如表 5 所示。从实验结果可以看出,本文提出的推理策略,在保证精度的同时,降低了冗余检测框的数量。

表 5　Pascal 2007 的检测结果

算法	主干网络	box-d	*mAP*/%
NMS	改进 VoVNet	2010	69.9
本文算法	改进 VoVNet	2002	70.4

图 7 是表 5 中本文算法分别对数据集和现实采样图片的可视化结果。可以看出,图 7 中检测框上标记的置信分数与图 5 相比发生了明显的变化,均比图 5 的数值大,说明类内分数重分配推理策略已经通过重叠检测框的 IoU 值再次计算了置信度,使得算法在确保每个对象对应一个检测框的同时,减少了冗余的检测框数量。

图 7　检测可视化结果

为了验证推理策略的实用性,在 FCOS 算法和已经训练好的 Faster RCNN 和 YOLO v3 算法上进行

验证,实验结果如表 6 所示。该实验是为了验证推策略的通用性,所以 3 个算法不同的训练设定不影响推理策略的有效性。表中第 1 列中的本文算法表示在算法的推理阶段使用的推理策略是本文提出的类内交互比分数重分配机制。3 组实验的实验指标均表明本文的推理策略可以减少检测框的冗余数量,从而提高了算法精度。

表 6　在 Pascal 2007 的检测结果

算法	主干网络	box-d	*mAP*/%
FCOS(NMS)	VoVNet v1-57 + Res	11 036	67.3
FCOS(本文算法)	VoVNet v1-57 + Res	**7590**	**68.1**
Faster RCNN(NMS)	VGG 16	3916	71.75
Faster RCNN(本文算法)	VGG 16	**3068**	**72.33**
YOLO v3(NMS)	EfficientNet	246	73.12
YOLO v3(本文算法)	EfficientNet	**109**	**73.18**

4　结　论

为了减少冗余检测框的数量,同时有效提升目标检测的定位精度,本文基于 RetinaNet 网络结构提出了一种减少类内检测框重叠推理策略的目标检测算法。实验结果表明,在选取的正样本中含有错误的类别标定或是由低质量样本预测得到低质量的检测结果,可以通过中心采样和类内分数的重分配影响检测框的 NMS 操作来减少错误的检测结果,从而提高检测精度。在实验过程中,本文算法仍然可以满足实时性的要求,今后的工作中还可以继续优化。除此之外,算法的参数量、简化网络结构等方面还需进一步改进。

参考文献

[1] 杨绪兵, 葛彦齐, 张福全, 等. 基于矩阵模式的林火图像半监督学习算法[J]. 图学学报, 2019, 40(5): 835-842

[2] 董正天, 刘斌, 胡春海, 等. 基于机器视觉的丝印样板表面缺陷检测方法研究[J]. 高技术通讯, 2020, 30(12):1309-1316

[3] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards realtime object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017,39(6):1137-1149

[4] LAW H, DENG J. CornerNet: detecting objects as paired keypoints[C]// Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany,2018: 734-750

[5] DUAN K, BAI S, XIE L, et al. Centernet: keypoint triplets for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 6569-6578

[6] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 9627-9636

[7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017:2999-3007

[8] LEE Y, HWANG J, LEE S, et al. An Energy and GPU-computation efficient backbone network for real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, USA, 2019: 752-760

[9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2017: 770-778

[10] WEI L, DRAGOMIR A, DUMITRU E, et al. SSD: single shot multiBox detector[EB/OL]. https://arxiv.org/abs/1512.02325:arXiv,(2015-12-08),[2021-05-19]

[11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2117-2125

[12] ZHU C, HE Y, SAWIDES M. Feature selective anchor-free module for single-shot object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019:840-849

[13] ZHANG S, CHI C, YAO Y, et al. Bridge the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2020: 9759-9768

[14] QIU H, MA Y, LI Z, et al. Borderdet: border feature for

dense object detection[C]// European Conference on Computer Vision, Glasgow, UK, 2020: 549-564

[15] NEUBECK A, GOOL L J V. Efficient non-maximum suppression[C]// International Conference on Pattern Recognition, IEEE Computer Society, Hong Kong, China, 2006: 1051-4651

[16] LIU Z K, HU J, WENG L, et al. Rotated region based CNN for ship detection[C]// IEEE International Conference on Image Processing, Beijing, China, 2017: 2381-8549

[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132-7141

[18] YU J, JIANG Y, WANG Z, et al. UnitBox: an advanced object detection network[C]// Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, Netherlands, 2016: 516-520

[19] EVERINGHAM M, WINN J. The PASCAL visual object classes challenge 2007 (VOC2007) development kit[J]. *International Journal of Computer Vision*, 2006, 111(1): 98-136

Object detection algorithm for reducing the number of redundant detection bounding boxes

WANG Xianbao, WU Menglan, YAO Minghai

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

Abstract

In the field of machine vision, many anchor-free object detection algorithms produce redundant bounding boxes when processing dense images, which reduces the detection precision. In view of this phenomenon, a target detection method is proposed which can reduce the number of redundant detection frames by using the network structure of RetinaNet. Firstly, in the feature extraction stage, a new attention mechanism is added to improve the expression ability of features. Then, in order to reduce the possibility of wrong label calibration in the positive samples, the position of the selected positive samples is filtered, and then the positive samples selected by the algorithm are input into the prediction branch to obtain the coordinates and confidence of the object bounding box. Finally, according to the location and classification results of the object bounding boxes, a new strategy of intersection and fraction redistribution is proposed, which can reduce the number of overlapping detection boxes and improve the precision of the algorithm. The effectiveness of this algorithm is verified on the open image data set. The results show that the proposed algorithm can improve the detection precision and optimize the positioning effect, and has a good application prospect.

Key words: anchor-free, object detection, attention mechanism, score redistribution, inference strategy